Virtual Reality exposure therapy for public speaking anxiety in routine care: a single-subject effectiveness trial

Philip Lindner, Jesper Dagöö, William Hamilton, Alexander Miloff, Gerhard Andersson, Andreas Schill & Per Carlbring

To cite this article: Philip Lindner, Jesper Dagöö, William Hamilton, Alexander Miloff, Gerhard Andersson, Andreas Schill & Per Carlbring (2020): Virtual Reality exposure therapy for public speaking anxiety in routine care: a single-subject effectiveness trial, Cognitive Behaviour Therapy, DOI: 10.1080/16506073.2020.1795240

To link to this article: https://doi.org/10.1080/16506073.2020.1795240
Virtual Reality exposure therapy for public speaking anxiety in routine care: a single-subject effectiveness trial

Philip Lindner, Jesper Dagöö, William Hamilton, Alexander Miloff, Gerhard Andersson, Andreas Schill and Per Carlbring

Department of Psychology, Stockholm University, Stockholm, Sweden; Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, & Stockholm Health Care Services, Region Stockholm, Stockholm, Sweden; PBM, Stockholm, Sweden; Mimerse, Stockholm, Sweden; Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

ABSTRACT
Virtual Reality (VR) can be used as a therapeutic tool to conduct efficacious in-session exposure therapy by presenting virtual equivalents of phobic stimuli, yet past hardware restrictions hindered implementation in routine care and effectiveness studies. The current study examines the effectiveness of a VR-assisted treatment protocol for public speaking anxiety with demonstrated efficacy, this time in routine care, using affordable VR hardware. Participants (n = 23) were recruited via a private clinic and treated by one of four psychologists with only minimal VR-training. Using a single-subject design and dual-slope modeling (adjusting the treatment-onset slope for treatment effects), we found a significant, large decrease in self-rated public speaking anxiety following the primary three-hour session, similar in magnitude to the previous efficacy trial. Multilevel modeling of in-session process measures suggests that the protocol works as intended, by decreasing catastrophic belief expectancy and distress, and increasing perceived performance quality. Adherence to the online transition program that followed—encouraging in-vivo exposure—was relatively poor, yet symptoms decrease continued. No change was observed over the three-month follow-up period. We conclude that VR exposure therapy can be effective under routine care conditions and is an attractive approach for future, large-scale implementation and effectiveness trials.

Introduction
Virtual Reality (VR) refers to a broad set of technologies that create an immersive experience of being present in a virtual, computer-generated environment (Botella et al., 2017). This is typically achieved using a head-mounted display that covers the user’s eyes with stereoscopic displays (thus withholding the outside world) that simulate depth perception and are interactive to head rotation, allowing the user to look around the virtual environment. Interactive audio is often used to increase immersion, and handheld controllers may be used as an additional way of interacting with the virtual environment.
environment, e.g. through use as a mouse pointer or virtual hands. Recognizable VR technology emerged already in the 1990s and has seen remarkable development in recent years. VR has several distinct advantages as a format for psychological interventions. The ability to create virtual equivalents of phobic stimuli, for example, solves many logistic issues in providing exposure therapy. More importantly, the complete control of the user’s visual and auditory input makes it possible to create therapeutic experiences not available in the real world, for example being able to customize stimuli material to the user’s idiosyncratic fears, or precisely manipulate the environment to target idiosyncratic catastrophic beliefs. Dozens of high-quality trials since the early 2000’s support the efficacy of VR exposure therapy (VRET) for anxiety disorders (Carl et al., 2019; Fodor et al., 2018; Opriş et al., 2012), showing effect sizes similar to in-vivo exposure therapy (Wechsler et al., 2019) and that treatment effects generalize also to reduced fear of real-world equivalent phobic stimuli (Morina et al., 2015). Low rates of deterioration have also been found (Fernández-Álvarez et al., 2019). Concerns about sample size and other methodological aspects have however been raised (Lanier et al., 2019; Page & Coxon, 2016), highlighting the need for more, rigorous clinical trials.

In particular, there have been few attempts to implement and evaluate VR-based treatments in routine care (Gega, 2017). Doing so may have several advantages: offering VR treatments may attract patients reluctant to seek traditional treatment (Garcia-Palacios et al., 2001) thereby addressing the large treatment gap for mental disorders (Wang et al., 2007); clinics may provide treatments alternatives that are not possible or practically infeasible or too costly to provide in other ways; and full or partial automation of treatment delivery (Donker et al., 2019; Freeman et al., 2018; Lindner et al., 2020) can be used to address difficulties in training clinicians to specific protocols and ensure a greater degree of standardization. Arguably, implementation was until recently hindered by the limitations of the previous generation of VR hardware, that was inaccessible, expensive (often costing more than 10,000 USD), required a high degree of technical skills to operate, were heavy to wear, and prone to induce motion sickness due to low graphical quality. The recent and continued release of consumer-targeted VR platform presents a paradigm shift in the overall accessibility and user-friendliness of VR, with high-quality standalone mobile headsets (requiring no smartphone or tethered gaming computer to run) costing as little as 240 USD (Lindner et al., 2017).

We have recently developed and evaluated a treatment protocol for public speaking anxiety (PSA) specifically for use with modern, off-the-shelf consumer VR technology, and which in a randomized controlled trial was associated with a large decrease in self-rated PSA (Lindner et al., 2019b). A non-randomized pilot study has recently replicated this treatment effect in adolescents using an age-adapted version of the protocol and custom stimuli material (Kahlon et al., 2019). The treatment protocol builds on traditional in-session exposure techniques (Ollendick & Davis, 2013; Öst, 1989), integrates novel therapeutic elements such as perspective-shifting audio playback (Nilsson & Lundh, 2016), and uses VR for stimuli presentation that would otherwise be logistically infeasible (i.e. requiring a hired, trained audience). PSA is a highly prevalent condition: one-third of the population reports excessive anxiety before speaking in front of an audience, and a further third reports clinically significant distress or interference with everyday life (Stein, 1996). Another study found PSA to be the most prevalent social fear amongst both individuals with social anxiety disorder (71.4%) and those without (8.7%)
(Tillfors & Furmark, 2007). The high prevalence and impact on many different life domains mean that clinicians may encounter this condition in a wide range of settings—from school and university health services, to occupational healthcare providers, primary care and in private practice—yet will face logistic barriers to providing effective exposure therapy. Previous research supports using VR for treating PSA (P. Anderson et al., 2003; P. L. Anderson et al., 2013; Harris et al., 2002; Hinojo-Lucena et al., 2020; Owens & Beidel, 2015; Parrish et al., 2016; Stupar-Rutenfrans et al., 2017; Wallach et al., 2009). A modern, evidence-based treatment protocol, specifically designed to make full use of consumer VR hardware, and packaged in a user-friendly application for stimuli presentation, would be of great clinical value. However, how well already established treatment efficacy translates into effectiveness when used in routine care remains to be examined. If VR is ever to become a common therapeutic tool, effectiveness studies are needed to demonstrate clinical utility also in routine care settings.

In the current study, using a single-subject design, we evaluate the effectiveness of a clinician-led, one-session VR exposure therapy (VRET) for PSA in a routine care setting, followed by an online transition program promoting subsequent in-vivo exposure. In addition, we evaluate whether the treatment works via the intended mechanisms using multilevel modeling of process measures collected in-session, exercise-by-exercise.

**Methods**

**Design, ethics and pre-registration**

This trial was pre-registered in the ClinicalTrials.gov registry prior to data collection (NCT03885414) and was approved by the Regional Ethical Review Board in Stockholm (2018/2657-32). Since the trial was conducted in routine care with enrolled patients, a waiting-list control group (entailing delayed treatment for study purposes) or individually randomized delay of treatment onset (a multiple baseline design) were deemed inappropriate. Instead, we opted for an AB single-subject design (Kazdin, 2018, December) that allows for each participant to serve as their own control by including multiple baseline and post-treatment assessments, and using piecewise regression with dual slopes to adjust the effect of treatment-onset for non-stable baseline trends (i.e. presumed continued spontaneous remission) (Moeyaert et al., 2014). Treatment was provided under the Swedish Health Care Act (not as research) by licensed clinicians in full compliance with required clinical procedure (full patient records kept, chain of care, etc.). In addition, the study was conducted in a routine care setting with regular clinic therapists providing treatment, minimal inclusion and exclusion criteria applied in recruitment, and with procedural flexibility extended as required, entailing that the current study meets criteria for an effectiveness trial (Zwarenstein & Treweek, 2009).

**Procedure**

Participants in the current study were either new or already enrolled patients at a local branch of PBM, a private chain of psychologist clinics specializing in providing cognitive behavior therapy for different mental and psychosomatic conditions. Patients come to PBM clinics
through either the public healthcare choice system (in select Swedish counties), occupational healthcare providers, or as self-referrals (covering their own treatment costs).

See Figure 1 for the flow chart. The study procedure closely followed the standard clinical procedure of PBM with some additions required for research purposes, primarily online outcome monitoring administered through the Iterapi platform (Vlaescu et al., 2016). Participants were recruited through one of two routes: already enrolled patients who had already completed the standard digital clinical assessment battery (not used in research), were informed about the possibility of VR exposure therapy for PSA by their initial psychologist, and referred to a PBM website for further information. This website in turn referred the patient to the study platform (Vlaescu et al., 2016), which included full study information on a public website, and allowed potential participants to register an anonymous account to provide informed consent and complete the study screening battery. Eligible participants were screened by PBM psychologists and contacted for a telephone interview, including a diagnostic interview (Sheehan et al., 1998) for social anxiety disorder and the performance-only specifier (Swedish MINI version 7 for DSM-5 criteria), before a final study inclusion decision was made by PBM staff according to the predefined inclusion and exclusion criteria (see below). Via the other route, new patients were recruited by PBM for this treatment and study specifically by informing existing corporate clients, and through media coverage and Google Ads. Like already enrolled patients, new potential patients were directed first to the PBM website, and then the study website where they completed the study screening battery and provided informed consent. Eligible patients then completed the same digital clinical assessment battery and

![Study flowchart](image-url)

**Figure 1.** Study flowchart.
telephone interview and were enrolled as regular patients. The primary outcome measure was administered once again approximately 1 and 2 weeks after completing the telephone interview.

The two recruitment routes then converged, with scheduling of a three-hour treatment session at one of two PBM clinics with a clinical psychologist. One week after the session, the primary and secondary outcome measures and additional forms were administered, after which a four week/module online transition program commenced via the same Iterapi platform (Vlaescu et al., 2016). During the transition program, the primary outcome measure was administered after the fourth and final week, followed by a 3-month follow-up assessment.

**Participants**

Recruitment to the study was done during a four-month window, with the aim of recruiting n = 25 participants, giving 80% power to detect a within-group effect size of d > 0.6 in a paired t-test, approximately d = 0.2 lower than that observed after the three-hour session in the previous study (Lindner et al., 2019b). Participants had to be 18 years or older, speak and understand sufficient Swedish to complete treatment, have stable internet access, and be able to travel to a PBM clinic on one occasion. The cost of treatment (3900 SEK, approximately 370 euros) would need to be covered either by the occupational healthcare provider or employer, or privately. Exclusion criteria included having a deficit in stereoscopic vision or problematic balance impacting the experience of VR, a severe psychiatric disorder that requires dedicated care (including severe major depression, alcohol or drug abuse, bipolarity, and psychosis), active psychoactive medication (unless stable for the last 3 months), or an ongoing or planned psychological treatment during the study duration. To be included, participants also had to score ≥60 on the Public Speaking Anxiety Scale (PSAS) (Bartholomay & Houlihan, 2016) serving as a primary outcome, corresponding to the 70th population percentile. Meeting criteria for social anxiety disorder was not an inclusion criterion, but all included participants did meet criteria, most of whom with the performance-only specifier. See Table 1 for participant characteristics.

<table>
<thead>
<tr>
<th>Measure</th>
<th>M or n</th>
<th>SD or %</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Female</td>
<td>n = 13</td>
<td>57%</td>
</tr>
<tr>
<td>Age</td>
<td>M = 40.61</td>
<td>10.15</td>
</tr>
<tr>
<td>% Married or committed relationship</td>
<td>n = 21</td>
<td>91%</td>
</tr>
<tr>
<td>% Completed university education</td>
<td>n = 17</td>
<td>74%</td>
</tr>
<tr>
<td>% Work as primary occupation</td>
<td>n = 20</td>
<td>87%</td>
</tr>
<tr>
<td>% Previous psychological treatment</td>
<td>n = 18</td>
<td>78%</td>
</tr>
<tr>
<td>% Previous psychoactive medication</td>
<td>n = 9</td>
<td>39%</td>
</tr>
<tr>
<td>% Self-referral</td>
<td>n = 22</td>
<td>96%</td>
</tr>
<tr>
<td>Impact on everyday functioning 0–10</td>
<td>M = 8.48</td>
<td>1.59</td>
</tr>
<tr>
<td>Self-rated duration (years) of clinically significant PSA</td>
<td>M = 19.26</td>
<td>9.93</td>
</tr>
<tr>
<td>Performance-only social anxiety disorder *</td>
<td>n = 16</td>
<td>70%</td>
</tr>
<tr>
<td>Time spent gaming each week (hours)</td>
<td>M = 1.13</td>
<td>2.60</td>
</tr>
<tr>
<td>% Any previous experience with VR</td>
<td>n = 7</td>
<td>30%</td>
</tr>
</tbody>
</table>

* All included participants met criteria for social anxiety disorder.
Treatment and therapists

Treatment closely followed the previously evaluated protocol (Kahlon et al., 2019; Lindner et al., 2019b). The protocol is grounded in a cognitive-behavioral model of PSA as a vicious circle of catastrophic beliefs about one’s own performance or appearance, or the views of others (Stein et al., 1996). This vicious circle leads to either complete avoidance of anxiety-provoking situations—entailing that catastrophic beliefs can never be falsified—or heightened physiological reactivity, self-attention and/or dysfunctional safety behaviors during endured anxiety-provoking situations. This in turn leads to poor performance and/or biased recall (Rapee & Hayman, 1996) that maintain and strengthen the underlying catastrophic beliefs. The treatment protocol targets specific catastrophic beliefs (Hofmann, 2007) through exposure exercises that emphasize inhibitory learning (Craske et al., 2014) over physiological habituation (A. Baker et al., 2010). This treatment rationale is made explicit to the patient (Bluett et al., 2016).

After an initial 15–25 minutes of psychoeducation, controlled and systematic exposure was conducted using a series of speech tasks (Hindo & Gonzalez-Prendes, 2011), both natural (e.g. presenting oneself) and artificial (e.g. quickly saying words beginning on a certain letter). All speeches are either fully improvised or performed after a short preparation time, with the rationale that extensive preparation serves as a safety behavior and that improvised speeches evoke more anxiety. While there is some progression in terms of difficulty, no idiosyncratic fear hierarchy was defined in order to promote variability, spontaneity, and sustained arousal (Culver et al., 2012). The therapist was encouraged to create custom speech tasks towards the end of the session, targeting idiosyncratic catastrophic beliefs. The initial six speech exercises were standardized, followed by up to five custom tasks all lasting 1–3 minutes each. The session ends with key treatment points and individual progression being summarized and continued practice in-vivo encouraged.

Unlike in the previous study, which used simpler, passive, 360 video-based stimuli material and the first generation of consumer VR hardware (Lindner et al., 2019b), the current study used a custom sandbox-type paradigm developed specifically for this treatment protocol, allowing users to build their own tailored exposure scenario by modifying parameters like setting and audience mood (see below). Using a user-friendly interface, the application allowed the user to customize the length of the task, audience mood (five sentiments), and three different environments (board room, conference room, and classroom) which varied with regards to audience size and proximity (Craske et al., 2014). The patient was encouraged to use this interface themselves, in order to strengthen their sense of control. The application ran on an affordable (240 euro), mobile VR platform, the Oculus Go.

Each task started with the therapist providing instructions on the speech task to be performed and under what circumstances, who then helped the patient to transform vague catastrophic beliefs into testable predictions. The patient was instructed to carry out each speech task as if it were a true speech. Once the environment was set up and the task performed in front of the virtual audience, the patient immediately rated the quality of their performance and their highest subjective units of distress (0–100). A selected set of audience behaviors (laughing, booing and applauding) could be manually triggered via buttons on the hand controller, either by the therapist or the patient. In the previous study (Lindner et al., 2019b), patients then listened to an audio recording of their speech while
using mental imagery to seat themselves as an audience member. This was done to aid disproving catastrophic beliefs through objective feedback. In the current study, the custom-built VR application automatically recorded audio and head rotation parameters during the task, and provided a playback option afterwards that allowed the user to see a humanoid avatar give the (lip-synced) speech as recorded, through the eyes of an audience member seated in the same environment that the speech was held. See Figure 2 for environment and playback screenshots. In addition to the examples described above, throughout the session, the therapist was tasked with providing clear instructions and tailoring the protocol, assisting in extracting testable hypothesis for the behavioral experiments through Socratic dialogue, highlight and discuss discrepancies, to monitor safety behaviors, reinforce any progress, and to steer away from any dysfunctional cognitions maintaining PSA but that are hard to test in VR (e.g. relating to audience perceptions) or
VR-specific safety behaviors (e.g. focusing on graphical inconsistencies in the graphical environment or light leaking in) (Lindner et al., 2020c).

One week after the 3-hour session, the 4-week/module online transition program commenced. The transition program repeated important psychoeducation from the session, yet focused on encouraging a transition to in-vivo exposure in everyday settings, providing planning and evaluation worksheets that the therapist could comment on.

Four licensed clinical psychologists, trained and experienced in cognitive-behavioral therapy, delivered treatment (including the transition program), with between two and seven patients each. Therapist allocation was quasi-random, i.e. no true randomness (no use of random sequence and allocation), but no systematic allocation either (e.g. no allocation of patients with severe symptoms to more experienced clinicians). Prior to participant enrollment, all therapists completed 3 hours of training in the specific treatment protocol and use of the VR hardware and software. Training took place in the form of a workshop held by a protocol author (PL), covering the underlying cognitive-behavioral model of PSA and derived treatment rationale, a set of points to cover in the introductory psychoeducation and final summary, along with demonstration and one-on-one supervised practice of how to conduct the exposure exercises using the VR hardware and software. Therapists were provided with a written treatment manual corresponding to the training session, yet no qualifying examinations, or further training sessions or method supervision sessions, were held.

**Measures**

**Primary outcome measure**

As in the two previous studies (Kahlon et al., 2019; Lindner et al., 2019b), the PSAS served as the primary outcome measure (the Swedish version). Seventeen items, covering cognitive, behavioral and physiological symptoms of PSA, are rated using a five-step Likert response format, giving a theoretical total score range of 17–85. In our sample of n = 23 included participants, internal consistency was calculated to α = .78 at screening (one item omitted in this specific analysis due to zero variance). The PSAS was administered a total of eight times during the treatment duration and once again at the three-month follow-up.

**Secondary outcome measures**

Secondary outcomes were measured at screening, post-session, after the transition period, and at the three-month follow-up. The Liebowitz Social Anxiety Scale Self-Rated (LSAS-SR) (S. L. Baker et al., 2002) and the Brief Fear of Negative Evaluation (B-FNE) (Carleton et al., 2011) were included to measure the broader clinical presentation of social anxiety. The Patient Health Questionnaire 9-item (PHQ-9) (Kroenke et al., 2001) and Generalized Anxiety Disorder 7-item (GAD-7) (Spitzer et al., 2006) were included to measure depressive and anxiety symptoms. Finally, the Brunnsviken Brief Quality of life scale (BBQ) (Lindner et al., 2016) was used to measure subjective quality of life. Established Swedish versions of all scales were used (Hedman et al., 2010; Johansson et al., 2013; Mörtberg & Andersson, 2013).
**Negative effects**

The Negative Effects Questionnaire (Rozental et al., 2016) was administered at the final post-assessment to screen for 32 possible negative effects of psychological treatment (Rozental et al., 2014). Of note, since this was after the transition program, any negative effect could be related to either the three-hour session or the transition program.

**In-session process measures**

Throughout the three-hour session, the therapist recorded several presumed process measures using a standardized session protocol, both prior to and after each speech exercise. This included ratings on self-rated quality of performance (immediately after exercise and again after seeing playback), subjective units of distress (rated immediately after the exercise, both maximum during and at the end), and catastrophic belief expectancy (prior to exercise and after seeing playback). All measures were rated on a scale of 0–100, with different but standardized verbal anchors at 0 and 100. Due to administrative errors (the source of which could not be reliably ascertained), in-session data from n = 17 patients were available for analyses. In cases of multiple numbers or ranges (e.g. “40 to 50”) written in cells, the first or highest number (respectively) was retained, regardless of whether it was pre or post an exercise. Since only the first six exercises were standardized, only data from these were retained for analyses, in order to calculate meaningful summary statistics across individuals. In total, k = 478 ratings were available for analyses.

**Analyses**

All analyses were carried out using the R statistical environment, using the *lme4* and *lmerTest* packages (Bates et al., 2015; Kuznetsova et al., 2017) for mixed effects modeling. Change in the primary outcome measure was modeled according to the intention to treat principle, using a dual-slope, piecewise mixed effects model estimating the change in slope after treatment onset (thus adjusting this slope for presumed continued pretreatment slope) (Hesser, 2015). Missing data were assumed missing at random, estimated using maximum likelihood and random effects modeling. Step-wise effect sizes were calculated on observed data under the missing at random assumption. In order to achieve meaningful equidistance when using a linear time predictor, we opted to use the same time coding scheme as in the previous study (Lindner et al., 2019b): the incremental value of the weekly measures during the transition period was set to Δ = 0.25 to account for hypothesized smaller effects than the 3-hour session. Other incremental values for the time variable were set to Δ = 1. See Table 2 for coding.

Since there were no multiple baseline measures of the secondary outcome measures over the course of the trial, these were modeled using the same coding scheme but with a single time predictor; thus, these parameter estimates are unadjusted. Since only one follow-up assessment was available (for which there was no obvious linear equidistance value) and attrition was high, a separate mixed model (random intercept only, no slope) was calculated for the primary outcome where both the last weekly measure and the post-transition assessment were time-coded as 0 and the follow-up as 1, in order to maximize
Table 2. Observed means, standard deviations and n for primary outcome (PSAS).

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
<th>Description</th>
<th>M</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>Screening</td>
<td>74.9</td>
<td>6.28</td>
<td>23</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>Pre: prior to interview</td>
<td>71.5</td>
<td>7.75</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>Pre: after interview</td>
<td>70.8</td>
<td>8.33</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>Post-OST</td>
<td>63.9</td>
<td>8.82</td>
<td>16</td>
</tr>
<tr>
<td>3.25</td>
<td>1.25</td>
<td>Week 1 of transition program</td>
<td>60.6</td>
<td>6.69</td>
<td>14</td>
</tr>
<tr>
<td>3.5</td>
<td>1.5</td>
<td>Week 2 of transition program</td>
<td>60.8</td>
<td>10.9</td>
<td>12</td>
</tr>
<tr>
<td>3.75</td>
<td>2.5</td>
<td>Week 3 of transition program</td>
<td>60.8</td>
<td>14.3</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>Final (post-transition program)</td>
<td>56.2</td>
<td>14.1</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Three-month follow-up</td>
<td>53</td>
<td>14.5</td>
<td>8</td>
</tr>
</tbody>
</table>

OST: One-session treatment.

power. For follow-up modeling of secondary outcomes, only the post-transition and final assessments were available, resulting in different number of data points.

In-session process data across exercises were modeled using separate linear mixed effects models, all including a time variable corresponding to the pre-post effect of each exercise: either pre–post the entire exercise (catastrophic belief expectancy), maximum and final value during exposure (SUDs), or pre-post seeing the playback (performance quality rating). For catastrophic belief expectancy and SUDs, a linear variable corresponding to exercise was also included; for quality, an obvious non-linear trajectory made this unsuitable.

Results

Attrition, program engagement and negative effects

A total of n = 23 (target n = 25) participants were included during the recruitment window of 117 days (January to May 2019). Of these, n = 20 completed the three-hour session, and n = 19 began the transition program (defined as opening the first module), with which engagement was generally low. See Table 3 for engagement statistics. The most commonly reported negative effects of treatment were “I felt I was under more stress” (n = 6), as expected from an exposure-based treatment, and “I felt that my expectations for the treatment were not fulfilled” (n = 6).

Primary outcome

Dual-slope mixed effects modeling revealed a significant change in slope after treatment-onset (B = −4.14, SE = 1.48, p = .0063), after an initial non-significant negative pre-treatment slope (B = −1.95 SE = 1.01, p = .057). Analyzing the treatment steps separately revealed an unadjusted observed effect size of the three-hour session of d = 0.77 growing to d = 1.15 after the transition period. See Figure 3 and Table 2.

Table 3. Participant engagement with the online transition program.

<table>
<thead>
<tr>
<th>Measure</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logins on platform</td>
<td>5.16</td>
<td>3.37</td>
</tr>
<tr>
<td>Opened modules</td>
<td>2.21</td>
<td>1.13</td>
</tr>
<tr>
<td>Incomplete in-vivo exercises</td>
<td>1.91</td>
<td>0.83</td>
</tr>
<tr>
<td>Complete in-vivo exercise</td>
<td>4.50</td>
<td>3.70</td>
</tr>
</tbody>
</table>
Secondary outcomes

Mixed effects models with a raw time variable revealed significant time effects on the LSAS-SR (B = −7.95, SE = 2.06, p = .0011), B-FNE (B = −1.59, SE = 0.72, p = .036) and GAD-7 (B = −1.08, SE = 0.49, p = .049), but not on the BBQ (B = 1.37, SE = 2.07, p = .52) or PHQ-9 (B = −0.78, SE = 0.433, p = .095). See Figure 4 for plots (including follow-up results not included in the above models).

In-session process measures

Mixed effects modeling of in-session process measures revealed a significant effect of completing an exercise on catastrophic belief expectancy (B = −21.55, SE = 4.48, p < .001),
SUDs (B = −19.65, SE = 2.9, p < .001), and quality (B = 11.1, SE = 3.1, p < .001). Ratings of catastrophic belief expectancy decreased linearly across exercises (B = −1.79, SE = 0.75, p = .019) while SUDs did not (B = −0.15, SE = 0.8, p = .85). See Figure 5.

**Long-term effects**

Random-intercept modeling of the follow-up period revealed no difference in PSAS scores over the three-month follow-up period (B = −1.05, SE = 2.32, p = .66, k = 29), or any of the secondary outcomes (p > .085, k = 20). These analyses are low-powered and should be interpreted with caution.

**Discussion**

Decades of research have revealed VR exposure therapy to be an efficacious treatment of anxiety disorders. With the advent of accessible, user-friendly consumer VR technology, these treatments are now ready to be implemented in regular care (Gega, 2017), yet must first be subject to the same scientific evaluation as other clinical interventions. To our knowledge, the current study is the first to examine the effectiveness of VR exposure therapy in a routine care setting, delivered by ordinary cognitive-behavioral therapists with no previous experience of using VR clinically. Observed effect sizes on self-reported PSA were remarkably similar to those reported in a previous efficacy trial of the same

---

**Figure 5.** Modeling of in-session process measures. Abbreviations used: CBE, catastrophic belief expectancy; SUDs, subjective units of distress. All measures rated 0–100.

SUDs (B = −19.65, SE = 2.9, p < .001), and quality (B = 11.1, SE = 3.1, p < .001). Ratings of catastrophic belief expectancy decreased linearly across exercises (B = −1.79, SE = 0.75, p = .019) while SUDs did not (B = −0.15, SE = 0.8, p = .85). See Figure 5.

**Long-term effects**

Random-intercept modeling of the follow-up period revealed no difference in PSAS scores over the three-month follow-up period (B = −1.05, SE = 2.32, p = .66, k = 29), or any of the secondary outcomes (p > .085, k = 20). These analyses are low-powered and should be interpreted with caution.

**Discussion**

Decades of research have revealed VR exposure therapy to be an efficacious treatment of anxiety disorders. With the advent of accessible, user-friendly consumer VR technology, these treatments are now ready to be implemented in regular care (Gega, 2017), yet must first be subject to the same scientific evaluation as other clinical interventions. To our knowledge, the current study is the first to examine the effectiveness of VR exposure therapy in a routine care setting, delivered by ordinary cognitive-behavioral therapists with no previous experience of using VR clinically. Observed effect sizes on self-reported PSA were remarkably similar to those reported in a previous efficacy trial of the same
protocol (Lindner et al., 2019b) and in a protocol adapted for adolescents (Kahlon et al., 2019), suggesting no loss of treatment effects in translating the intervention into routine care. Research conducted both before (Schwartzman et al., 2012; Segal et al., 2011) and after the advent of consumer VR technology (Lindner et al., 2019d) has revealed that clinicians have favorable attitudes towards VR, yet show some concerns with financial costs and usability. The current study suggests that any such concerns can be mitigated by offering ordinary clinicians a relatively short training program, and user-friendly VR software and hardware.

In addition to demonstrating effectiveness, the current study also examined whether the treatment protocol works through the intended mechanisms, by statistical modelling of in-session process measures. Although no proper mediation analysis can be performed since there was no control group (Hesser, 2015), multilevel modeling of in-session measures revealed that the VR exposure exercises did indeed lower catastrophic belief expectancy and SUDs, and that subjective performance quality ratings increased after seeing an innovative perspective-shifting playback that allowed the user to see an avatar holding a record of their speech while now they themselves sitting in the audience. The treatment protocol, including exercise design, was explicitly developed with an inhibitory learning rationale and did not emphasize or require habituation. In addition to the wealth of experimental support for this approach over habituation (Craske et al., 2014) —although challenged by recent research (Scheveneels et al., 2019)—this decision was also based on a hypothesis, supported by the physiological findings from the pilot replication study in adolescents (Kahlon et al., 2019), that the VR public speaking paradigm would not evoke a physiological response on par with real-life public speaking, even among users with severe PSA. While previous research has shown that VR public speaking does evoke a significant physiological and subjective fear response, this response is small in terms of absolute numbers (Owens & Beidel, 2015; Takac et al., 2019) and lower than that observed in for example PTSD (Rauch et al., 2018). Arguably, this contrast is due both to the complexity of presenting human stimuli from a graphical perspective (Seyama & Nagayama, 2007), and that many of the most common catastrophic beliefs related to public speaking concern perception by others (Stein et al., 1996), which may be hard to evoke when the user knows that the audience are not real people.

As in the previous study, immediate treatment results generalized to the broader clinical presentation of social anxiety (as measured by the LSAS) and fear of negative evaluation (as measured by the B-FNE), but not depression, or quality of life. Pre-treatment scores on the latter two measures were however not in the clinical range; hence, significant changes were not expected. The significant but weak decrease in fear of negative evaluation found in the current study supports the hypothesis that VRET for PSA works primarily by disproving catastrophic beliefs about one’s own performance and display of physiological symptoms, rather than how one is perceived by others. Congruently, meta-analytic research has revealed that while effect sizes for VRET and in-vivo exposure therapy do not differ for specific phobia and agoraphobia, the effect size for social phobia significantly favored in-vivo (Wechsler et al., 2019). This analysis did however include studies that featured exposure for not only performance situations like public speaking, but also social interactions, the presence of which is associated with a different, more severe clinical presentation (Blöte et al.,
2009). In the current study, we refrained from correlating different process measures with outcomes due to the low number of participants. Future research on the mechanisms of VRET for PSA should include larger samples and appropriate comparison groups that allow statistical mediation modeling of presumed processes on outcomes, and examine whether VR-specific safety behaviors (such as stressing to oneself that the audience is not real) are associated with reduced fear response and treatment effects. Additionally, the sandbox mode used in the current study that allowed tailoring of the exposure scenarios—an inherent advantage of VRET (Rizzo et al., 2010)—could in theory also be used as a safety behavior akin to e.g. PTSD patients withholding particularly traumatic aspects of a trauma to avoid exposing themselves to this particular aspect. Future research should investigate the prevalence of this, and also to test the added benefit of tailored over non-tailored VR exposure scenarios in terms of efficacy. One recent study showed that exposure to standardized VR worry scenarios is equally effective in reducing anxiety as exposure to personalized imagined scenarios (Guitard et al., 2019). Further, most past VRET trials have had limited tailoring capabilities yet still report large effect sizes (Carl et al., 2019; Fodor et al., 2018; Opriş et al., 2012), suggesting that tailoring is not a requirement for effect, although this does not exclude the possibility of larger effects with tailored stimuli material. An additional avenue of future research is examining whether exposure using so-called social VR (Oh et al., 2018), that allow virtual gatherings through avatar representation of multiple users, evokes a stronger fear response, and whether the same effect can be achieved by blinding the patient to whether the audience members are real people represented as avatars, or so-called non-playable characters.

The results of this first effectiveness trial (and previous studies) show that VR can be successfully implemented in routine care; the next translational step would thus be larger, randomized effectiveness trials over longer periods of time. Here, the aim should be two-fold: both demonstrating non-inferiority when compared to gold-standard in-vivo equivalents (Miloff et al., 2019), but also superiority against treatment-as-usual. The latter is important since one of the primary clinical advantages of VR is that it solves logistic issues in providing certain types of treatment which may not otherwise be provided at all; thus, comparisons against common treatment-as-usual alternatives (e.g. natural waiting-list for group treatment of PSA) would serve to inform clinics on the benefits of using VR in a wider perspective.

The comparatively lower adherence to the online transition program observed in the current study, primarily in terms of module engagement, is not unexpected given that adherence is typically lower in effectiveness than efficacy studies. In addition, several study-specific aspects may also explain this finding: there is no way of ascertaining to what degree each participant was informed about the importance of continued training; the clinical circumstances may have hindered therapists from being more active on the online treatment platform (e.g. not sending enough personal reminders); and since treatment was associated with a certain cost (paid for by either public healthcare, the occupational healthcare provider, or personally), patients may have felt more free to deviate from adherence recommendations. The relatively small sample size and lack of a control group preclude us from using more advanced statistical techniques like Complier Average Casual Estimation (Hesser et al., 2017) to properly estimate the effects of adherence to the transition program. Further symptom reduction during the transition phase was lower than in the efficacy study (Lindner et al., 2019b), congruent with the hypothesized effect of performing in-vivo exposure tasks.
Research on the transition from VRET to in-vivo exposure, and associated impact on further symptom reduction, is almost non-existent and should be considered an important topic for future research. Two findings from the extant literature are however indicative of causal effects of subsequent in-vivo exposure after VRET on further symptom reduction. First, in the therapist-led arm in the efficacy trial, time-lagged multilevel modeling revealed significant additional symptom decrease in participants with at least one logged in-vivo exercise a particular week during the transition phase (Lindner et al., 2019b). Second, two studies on the same automated VRET intervention for spider phobia differed in long-term effects: in the study that emphasized transitioning to in-vivo exposure, symptom reduction continued after VRET (Miloff et al., 2019), while in the study that did not discuss such a transition at all, no further symptom reduction was observed (Lindner et al., 2020). Since none of these findings definitively reveal a causal role, future trials should feature a secondary randomized allocation to either a promoting in-vivo exposure or not, after the same VRET intervention.

In addition to using VR as a therapist tool to conduct otherwise traditional psychotherapy (as in the current study), clinics also have the possibility of offering automated VR interventions, evidence-based versions of which are either already available for other anxiety disorders (Donker et al., 2019; Freeman et al., 2018; Miloff et al., 2019) and feasible for other disorders (Lindner et al., 2019a, 2019c). Having patients engage with automated VR treatments at a physical clinic, or being sent a VR headset for use at home (Donker et al., 2019; Lindner et al., 2019b), would resolve the arguably greatest obstacle to disseminating VR interventions at present, the still comparably low degree of adoption by consumers. Similar procedures already exist for the distribution of other medical devices, yet realistically, healthcare reimbursement systems (public, insurance or occupational) would however need to accommodate this treatment modality before implementation of this approach can be pursued.

**Limitations**

This first effectiveness study on VRET in routine care has some limitations in need of recognition. First, modeling change within-subject rather than between-subject entails a lower degree of certainty in any causal claims. Also, while spontaneous remission during the baseline phase was not unexpected and statistically adjusted for using duals-slope modeling, a stable baseline would have been preferable. The lack of a comparison group also means that we cannot statistically test whether change in catastrophic belief expectancy, SUDs or perceived performance quality mediates symptom reduction. Further, the study included no repeated measures on the content of the reported catastrophic beliefs, meaning that there is no way to evaluate which domains of catastrophic beliefs that could be manipulated through treatment. Only self-reported distress, and no psychophysiology, was measured in-session. As is typical in effectiveness trials (Zwarenstein & Treweek, 2009), protocol adherence by a therapist was not measured in order to mimic real-life circumstances. The number of treated participants per therapist was too low to reliably estimate therapist effects (Magnusson et al., 2018), although the use of several therapists should be considered a strength of an effectiveness study. Neither did the study feature a design allowing disentangling the effects of specific treatment components. Missing data at follow-up entails that estimates of long-term effects are low-powered. These limitations should be addressed in future trials.
Conclusions

One-session Virtual Reality exposure therapy can be an effective treatment of public speaking anxiety, also when delivered under routine care conditions by cognitive behavior therapists with no previous clinical experience of VR and only minimal training. Multilevel modeling of in-session process measures suggests that the treatment works through the intended mechanisms by decreasing catastrophic belief expectancy, increasing self-rated quality of performance and within-task habituation.

Acknowledgments

The authors would like to thank all patients for their participation; clinical psychologists Loo Westfelt, Maria Engman for providing treatment; George Vlaescu for technical assistance with the Iterapi platform; and the VR content developers at Mimerse.

Disclosure statement

Author AS is employed by the private chain of clinics at which the research was conducted, which now offers the same treatment as part of regular care; JD was employed at this chain of clinics at time of data collection. Author WH is the founder and chief technology officer of Mimerse, the company that developed the VR application used in the study. Author PL has consulted for Mimerse, but holds no financial stake in the company and does not stand to gain financially from the publication of this manuscript. The other authors report no conflicts of interest related to the work described: principal investigator PC stands for the integrity of the data.

ORCID

Philip Lindner https://orcid.org/0000-0002-3061-501X
Jesper Dagöö https://orcid.org/0000-0001-5879-2168
William Hamilton https://orcid.org/0000-0003-0482-9521
Alexander Miloff https://orcid.org/0000-0002-9125-8060
Gerhard Andersson https://orcid.org/0000-0003-4753-6745
Andreas Schill https://orcid.org/0000-0002-5136-212X
Per Carlbring https://orcid.org/0000-0002-2172-8813

References


Guitard, T., Bouchard, S., Bélanger, C., & Berthiaume, M. (2019). Exposure to a standardized catastrophic scenario in virtual reality or a personalized scenario in imagination for


What kind of randomized trials do patients and clinicians need? Annals of Internal Medicine, 150(10), 735. https://doi.org/10.7326/0003-4819-150-10-200905190-02002